Computing and Informatics, Vol. 31, 2012, 225–244

# AN IMPLEMENTED APPROACH FOR POTENTIALLY BREAST CANCER DETECTION USING EXTRACTED FEATURES AND ARTIFICIAL NEURAL NETWORKS

Heba Al-Hiary, Basim Alhadidi

*Al-Balqa Applied University*
*Al-Salt, Jordan*
*e-mail:* hhiary78@yahoo.com, hadidi72@hotmail.com

Malik Braik

*University of Birmingham*
*United Kingdom*
*e-mail:* m_fjo@yahoo.com

Communicated by Vladimír Kvasnička

**Abstract.** Breast cancer (B-cancer) detection is still complex and challenging problem, and in that case, we propose and evaluate a four-step approach to segment and detect B-cancer disease. Studies show that relying on pure naked-eye observation of experts to detect such diseases can be prohibitively slow and inaccurate in some cases. Providing automatic, fast, and accurate image-processing-and artificial intelligence-based solutions for that task can be of great realistic significance. The presented approach itself scans the whole mammogram and performs filtering, segmentation, features extraction, and detection in a succession mode. The feasibility of the proposed approach was explored on 32 commonly virulent images, and the recognition rate achieved in the detection step is 100 %; further, the approach is able to give reliable results on distorted medical images, since the approach is subjected to a rectification step. Finally, this study is very effectual in decreasing mortality and increasing the quality of treatment of early onset of B-cancer.

**Keywords:** Extracted features, artificial neural networks, morphological operations, mammogram images

## 1 INTRODUCTION

Breast cancer (B-cancer) is the most frequent tumefaction in women and is the leading cause of cancer deaths among women. Most people turn away to think of B-cancer as a woman's disease. But men get B-cancer too. Research has shown that women with a family history of B-cancer have a higher hazard for evolving the disease. That is true whether the family history is on the mother's side or the father's. In spite of the increasing number of cancers being diagnosed recently, the death rate has been reduced obviously due to the advanced screening programs [1]. Premature detection of B-cancer increases the potentiality of survival rate whereas postponed diagnosis considerably encounters the patient to a critical stage and occasionally results in death [2]. Proper screening programs and diagnostic techniques dramatically increase the survival rate of diagnosed women. Mammography is a technique that uses X-rays to provide an image of the breast. These images, called mammograms, are used to find potential signs of B-cancer like tumors and abnormal changes in the skin. Contemporarily, screening mammography of the breast are the most effective tools for premature detection of B-cancer. As a result, the issue of adverse consequences of screening for women who do not have B-cancer, as well as women who have an early stage of B-cancer that will not progress, has become one of the core issues in recent debates about mammography [3]. Still, studies have proved that all B-cancers that are retrospectively detected on the mammograms are not exactly detected by radiologists [4, 3]. Due to the subtle and complex nature of the radiographic findings concerned with B-cancer, human factors such as distraction by image features and simple oversight can be responsible for the errors in radiological diagnosis [5, 6]. The digital mammograms have already been classified as normal, cancerous or benign. Figure 1 shows an example of a typical mammogram.

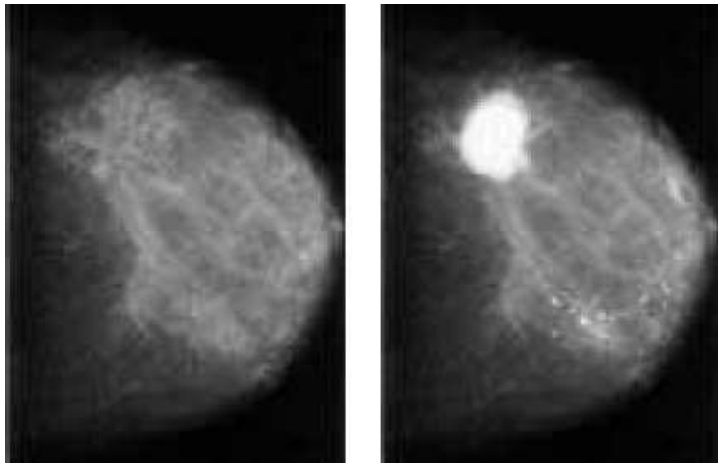

Fig. 1. Typical mammogram; left: Normal breast, right: Cancerous breast

This paper proposes a new algorithm addressed to perform prescreening of digital mammograms for the presence of virulent regions in breast tumors. Our algorithm partitions a mammogram image into homogeneous regions for possible locations of virulent regions in breast based on some morphological operations. Then, ANN is involved as a diagnostic technique depending upon some textural features composed of statistical gray level and Haralick features of the virulent regions. The performance achieved with the proposed scheme was 94.2 % sensitivity and 93.7 % specificity on test set. Results also show 100 % classification accuracy on training set of both sensitivity and specificity. Furthermore, promising results of both segmentation and detection phases were achieved. In conclusion, the aim of this work is threefold:

1. identifying region of interests in breast tumors;
2. analyzing the underlying features extracted from images of breast;
3. distinguishing between "cancer" and "non-cancer" images of B-cancer.

## 2 PROPOSED DETECTION ALGORITHM

Detection of abnormalities in mammograms is a difficult task, especially due to the high number of normal cases. Also, a closer inspection of mammograms reveals several difficulties for possible cancer detection, for example: the variations in the recording procedure as well as the global appearance may be both inefficient and difficult in detecting abnormalities in mammograms. A common approach for detecting abnormalities is to use a series of heuristics approaches, like filtering and thresholding. These methods may include texture analysis which may automatically detect abnormalities [7]. However, those methods possibly suffer from a lack of robustness when the number of images to be classified is large [8]. More than that, these approaches act only as a second reader and the final decision is made by the radiologist. By using Computer Aided Detection (CAD) software the number of errors might decrease, both false negatives (malignant cases that were not recalled) and false positives (cases that are recalled unnecessarily). Therefore, our work can contingently provide clues to treat B-cancer in its early stages. Precisely, interpreting mammograms that are used for diagnosing process involves preprocessing and post-processing stages. In this work preprocessing stage deals with image enhancement and image segmentation sub-stages. The post-processing stage consists of another two sub-stages; features extraction and neural network classifier. An overview of the basic steps describing the proposed approach is presented in the following steps:

1. mammogram image acquisition
2. perform image filtering using contrast stretching method
3. segment the filtered image based upon morphological operations
4. extract texture features and perform a texture statistics computation

5. detect the b-cancer status using neural network classifier tool
6. classify the status of the cancer as whether negative or positive.

The first step is the image filtering step which is very essential for premature detection of cancers. The first step in turns should be scanned for extracting the region of interest using the second step. The second step is then subjected to the third step in which some texture and statistical features are extracted. Finally, the extracted features (statistical gray level features and Haralick features) are passed through pre-trained neural networks for possible B-cancer detection. We illustrate the general outline and details of the presented work in the following sections.

## 3 IMAGE FILTERING

Medical images suffer from a wide variety of distortions which may cause some difficulties in identifying the abnormalities in breast tissues. As a result, perceptual quality of the images is deteriorated. High quality images and mammographic interpretation are mandatory for the detection of premature and delicate symptoms of B-cancer. Therefore, image filtering is inquired to suppress unwanted noise and improve both the quality and visual appearance of the images. The essential improvement that is required in mammography is accomplished by contrast stretching method, which is particularly beneficial for mammograms with low contrast.

### 3.1 Contrast Stretching

Contrast stretching is one of many techniques that used to enhance mammograms. It increases the contrast of the image, and adjusts the histogram of the image such that there is a greater separation between foreground and background gray level distributions [9]. Thus, an input mammogram is processed with contrast stretching method to obtain an enhanced image. The contrast filter is a spatial filter that has been developed to approximately match contrast variations of typical mammogram. Then, it provides an output with more contrast and detail. Consequently, at the locations of microcalcifications, the pixel values in the contrasted image are increased relative to the values of pixel of normal tissue [10]. Actually, contrast filter improves the readability of the low contrast areas in the image, and destroys the areas in the image where the intensity of the pixels is outside the range of intensities that are being enhanced [11, 12].

## 4 IMAGE SEGMENTATION

We need a significant technique for image segmentation that attempts to recognize and extract the malignant tumors with the aid of the distribution of gray levels in the image. So, one extreme task in discrimination the malignant tumors from benign tumors depends originally on choosing a proper segmentation process. Malignant tumors can have speculated and/or fuzzy borders, while benign tumors can

be lucent at the center and can have well defined borders. However, this step is developed because it can help radiologists identify the signs of cancer, and mark the suspicious areas in the mammograms, which may indicate the presence of cancer. There is a possibility for the difference in malignant tissues and normal tissues to exist in the mammograms but beyond the threshold of human perception. Therefore, we think deeply not to use thresholds in our proposed segmentation algorithm. The developed segmentation algorithm is based primarily on morphology theorem. Morphology algorithm is frequently employed in various areas to segment digital mammogram and is utilized in the construction of the regions. Once the tumor objects are extracted, these can be categorized as benign or malignant. Briefly, this step provides the following goals:

1. obtains the true locations of suspicious areas which may assist radiologists during the diagnosis
2. classifies the abnormalities of the mammograms into benign or malignant
3. spots the salient regions in mammograms; these salient regions correspond to distinctive areas which may include the breast boundary, the pectoral muscle, candidate masses, and some other dense tissue regions.

### 4.1 Morphological Operations

Morphological operations are a way of extracting image elements like regions shape, image boundaries and so on. Morphological theory is a powerful tool to analyze and describe numerous image processing problems [9]. Dilation and erosion are the two basic morphological operations. *Dilation* is usually used to smooth boundaries of the regions or bridge very small gaps between neighboring regions. *Erosion* has an opposite effect than dilation, as it shrinks the objects uniformly.

With $A$ and $S$ as sets in $Z^2$, the *dilation* of $A$ by $S$ is defined as given in Equation (1) [9].

$$A \oplus S = \{z|(\hat{S})_z \cap A \neq \Phi\} \tag{1}$$

With $A$ and $S$ as sets in $Z^2$, the *erosion* of $A$ by $S$ is defined as given in Equation (2) [9].

$$A \ominus S = \{z|(S)_z \subseteq A\} \tag{2}$$

In this work, $A$ is the image and $(S)$ is a matrix that defines a neighborhood shape and size for morphological operations consisting of only 0's and 1's. The pixels with values of 1 define the neighborhood. Experimentally, in this work, $(S)$ has a disk shape with radius equal to 5. The proposed methodology of segmenting B-cancer based on morphological operations can be concisely summarized into a group of steps as follows:

**Step 1:** The Region Of Interests (ROIs) is selected for the purpose of segmentation. ROIs is then identified and labeled. ROIs could be tumor or any other relevant matter.

**Step 2:** After specifying the ROIs, all small undesirable objects are reduced or completely removed.

**Step 3:** After eliminating all tiny objects, the image is smoothed using multidimensional filter.

**Step 4:** Dilation and erosion are applied using flat linear structuring element in each specified neighborhood for the selected ROIs in Step 3.

**Step 5:** Suppressing light structures that are connected to image border to reduce the overall intensity of the ROIs.

**Step 6:** Reconstructing the final segmented image from the previous two steps.

## 5 FEATURE EXTRACTION

From each segmented image, I, we extract a set of statistical- and texture-based features to capture the discriminating characteristics of the tissue patterns. A feature vector $(f)$ is created for I where each element of $(f)$ is a distinct feature value. The feature vectors in our study were extracted from the selected regions of the breast. These values are calculated as described below.

### 5.1 Textural Features

The proliferation of nuclei in cancerous tissue suggests that textural characteristics can help discriminate between different grades of B-cancers [13]. The following features are extracted from the 2D grayscale mammogram image.

#### 5.1.1 Grey Level Features

We calculate two gray level features from I as described in [14]. These features are:

**Average (mean):** mean $(m)$ measures the average intensity. Mean is defined as given in Equation (3) [15]

$$m = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{3}$$

where $x_i$ is the value of the pixel $i$, and $n$ is the total number of pixels.

**Standard Deviation:** standard deviation $(\sigma)$ measures the pixel deviation from the mean. $(\sigma)$ is defined as given in Equation (4) [15].

$$\sigma = \left[\frac{1}{n-1}\sum_{i=1}^{n}(x_i - m)^2\right]^{\frac{1}{2}} \tag{4}$$

#### 5.1.2 Haralick Features

The second-order co-occurrence texture features are described by sixteen Haralick features presented in [16]. In this study we calculate a co-occurrence matrix $Z \in R$

for image I, then we used the matrix to generate five Haralick texture features. Concisely, the method followed for extracting the feature set is called the Color Co-occurrence Method or CCM method in short. CCM is a method, in which both the color and texture of an image are taken into account, to arrive at unique features, which represent that image.

### 5.2 Co-occurrence Methodology for Texture Analysis

The image analysis technique selected for this study was the CCM method. The CCM methodology was applied to each pixel in the segemented image, in which each pixel map is used to generate a color co-occurrence matrix, resulting in one CCM matrix. The color co-occurrence texture analysis method was developed through the use of Spatial Gray-level Dependence Matrices or in short SGDM's [16]. The gray level co-occurrence methodology is a statistical way to describe shape by statistically sampling the way certain grey-levels occur in relation to other grey-levels. These matrices measure the probability that a pixel at one particular gray level will occur at a distinct distance and orientation from any pixel given that pixel has a second particular gray level. For a position operator $p$, we can define a matrix $P_{ij}$ that counts the number of times a pixel with grey-level $i$ occurs at position $p$ from a pixel with grey-level $j$. The SGDMs are represented by the function $P(i, j, d, \theta)$ where $i$ represents the gray level of the location $(x, y)$ in the image $I(x, y)$, and $j$ represents the gray level of the pixel at a distance $d$ from location $(x, y)$ at an orientation angle of $\theta$. All the neighbors from 1 to 8 are numbered in a clockwise direction. Neighbors 1 and 5 are located on the same plane at a distance of 1 and an orientation of 0 degrees. In this research, a one pixel offset distance and a zero degree orientation angle was used. We used GLCM function provided by Mathworks to create gray-level co-occurrence matrix for the images produced from the segmentation phase. The images are in a gray-level representation, with the number of gray levels set to 8, the symmetric value set to "true", and the offset value "0". Consequently, the statistics of the two-dimensional (2D) co-occurrence matrix of an image form the basis of our proposed technique for detection.

### 5.3 Texture Features Identification

Haralick [16] proposed sixteen features extraction approach from an image focusing on the spatial distribution of gray values. We prefer five features to complete this work, Those features are:

**Contrast:** contrast ($C$) measures the local variations in the gray level of the image. Contrast is defined as given in Equation (5) [15, 17]

$$C = \sum_{i=1}^{n} \sum_{j=1}^{n} |i - j|^2 \times P(i, j) \tag{5}$$

where $P(i, j)$ is the pixel at $(i, j)$ coordinates.

**Energy:** energy ($E$) measures the image uniformity. Uniformity measures the maximum gray level when they are equal. $E$ is defined as given in Equation (6) [15, 17].

$$E = \sum_{i=1}^{n} \sum_{j=1}^{n} P(i,j)^2 \tag{6}$$

**Dissimilarity:** dissimilarity ($D$) measures the images local variation along a certain orientation and displacement. $D$ is defined as given in Equation (7) [15, 17].

$$D = P(i,j) * \mid i - j \mid \tag{7}$$

**Homogeneity:** homogeneity ($H$) measures the relative smoothness of the intensity in the region. $H$ is defined as given in Equation (8) [15, 17].

$$H = \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{P(i,j)}{1 + \mid i - j \mid} \tag{8}$$

**Entropy:** entropy ($e$) measures the randomness in the image. $e$ is defined as given in Equation (9) [15, 17].

$$e = -\sum_{i} \sum_{j} P(i,j) \log_2 P(i,j) \tag{9}$$

Finally, all the seven extracted features of the segmented objects are passed to the final step of proposed B-cancer detection algorithm.

## 6 CANCER DETECTION USING ANN

The detection step has been accomplished based on backpropagation ANN. There are many classifiers that an be used in image processing appliations, and therefore, careful consideration was given to this particular stage. However, a neural network is chosen in this paper due to its well known technique as a successful classifier for many medical applications [18]. The inputs of the ANN are features vectors consisting of the seven extrated features. Those extracted features fall into three categories that are related to the statistical approach in which they are depending on the pixel gray values, neighborhood of the pixel, and pixel location. The features vectors serve as a classification method. The specified ROIs objects are split into two parts; training and testing sets. Then, once the feature extraction step was completed, two files were obtained. They were:

1. training texture feature data, and
2. test texture feature data.

The training sets are used to train the NN model, whilst the testing sets are used to verify the accuracy of the trained NN model. Finally, ANN is trained several times

for the possible detection of B-cancer cells, after that, ANN can classify the specified ROIs if it has B-cancer or not, and hence the B-cancer object is reconstructed. After that, we formulated a Matlab program aiming to specify if the input image belonging to class 1 or class 0, i.e. if it contains B-cancer or not.

## 7 EXPERIMENTAL RESULTS

The experimental results of the implemented detection approach are displayed in the next sub-sections in a succession mode. We present the results using some malignant samples.

### 7.1 Filtering Using Contrast Stretching

Filtering is a crucial step for handling the distortions that may exist in the mammograms, such that the suspicious region ould be displayed with higher contrast, which in turn accomplishes two tasks: the former task facilitates the function of the segmentation step for manipulating the breast cells describing the ROIs, the latter task augments the clarity of the object features which in turn paves the way for NN to classify the B-cancer easily. The results of applying contrast stretching to an image containing B-cancer is shown in Figure 2.

It can be seen from Figure 2 that the malignant regions appear almost visibly. In addition, the overall intensity of the processed image is also increased.

### 7.2 Segmentation Using Morphological Approach

The segmentation step is used to select and identify the ROIs which contains the virulent cells, The procedure given in Section 4 is applied for the whole set of tested images. Figure 3 shows an example of an original image containing B-aner, and the results of the segementation procedure.

In most cases, mammograms involve several dark regions and noises; occasionally, those dark regions could not be removed completely in the filtering step, which may trick the segmentation process; besides this, some small objects may be produced as a result of the first step in the segmentation procedure. These small undesirable objects are shown in the upper right image in Figure 3. Step 2 and step 3 in the segmentation approach try to remove all the tiny protuberances appeared on the upper right image as shown in the lower right image in Figure 3. The B-cancer in the lower right image is not fully identified, therefore, the last three steps in the segmentation procedure are inescapable to strengthen the segmentation result as shown in the lower right image in Figure 3. The last three steps in the proposed morphological algorithm eliminate the remaining dark regions and tiny protrusions from the image using the proposed filters, namely multidimensional filter, erosion filter, and the dilation operator. The dilation operator is applied to bridge the small gaps using the proposed flat structuring element.

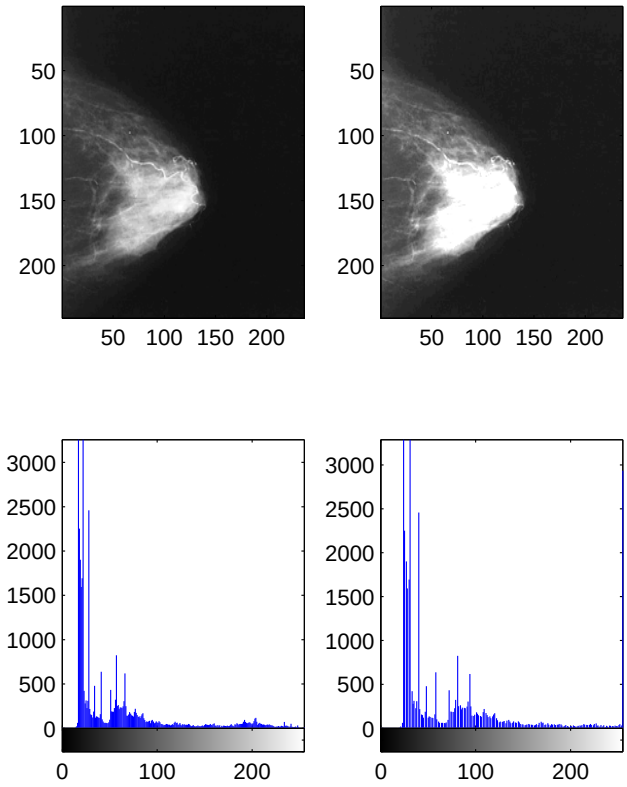


Fig. 2. Enhancement results: upper left: original image, upper right: enhanced image using contrast stretching, lower left: histogram of the original image, lower right: histogram of the stretched image

It can be seen from Figure 3 that the morphological operations perform very well in extracting the image elements, and it can express the image details in more specific way. As an additional step, the suspicious object in the input image is identified in the mammogram image by tracing the boundary of the suspicious regions. Through the proposed segmentation method, the binary image contains only 0's that represent the background and 1's that represent the object itself. In edge based techniques, segmenting an object is achieved by locating its boundary using image gradient, which has high values at the edges of the objects to locate the boundary of pixels, such that the edges in the image can occur on the boundary between two pixels when the respective brightness values between two pixels are significantly different [19, 5]. Hence, they are looking for edges between regions with different characteristics such that the edges are placed in the image with strong intensity contrast. The aim is to find object boundaries and segment regions enclosed by the boundaries. The edge

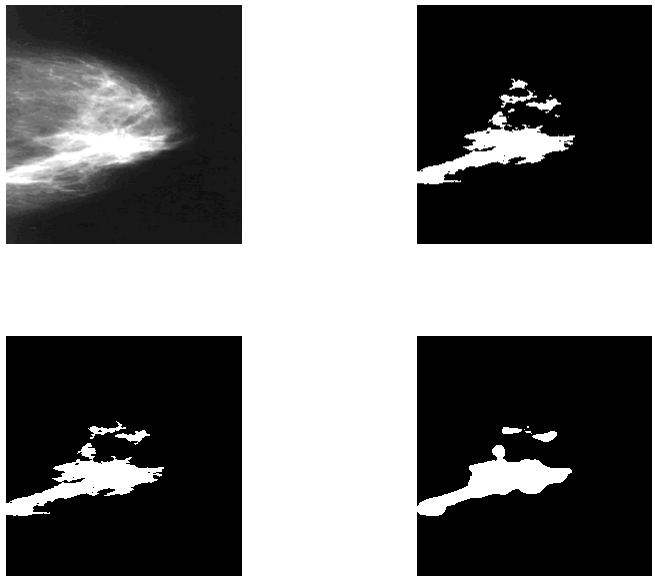

Fig. 3. Results of the proposed morphological approach: upper left: the enhanced image, upper right: the suspicious ROIs, lower left: eliminating all small objects from ROIs, lower right: the final segmented image

operator used in this work is Sobel edge detector [9]. Tracing the ROIs in the image containing B-cancer with the binary image and the edges of the cancer object are shown in Figure 4.

As a conclusion, the presented segmentation approach gives reliable results; because of that, morphological operations are extensively used in the analysis of medical images.

### 7.3 Extracted Features Using Co-Occurrence Methodology

We computed some of textural features with the aid of SGDM's using o-occurrence methodology; these features were used as the basis of our detection algorithm. So, they are considered for both train and test files of neural network.

### 7.4 Detection Using ANN

A software routine was formulated in MATLAB that would take `.mat` files representing the training and test data, train the classifier using the 'train files', and then use the 'test file' to perform the classification task on the test data. Consequently, the routine would load all the training data files and make modifications to the data according to the proposed model chosen. Anyhow, the detection step is performed

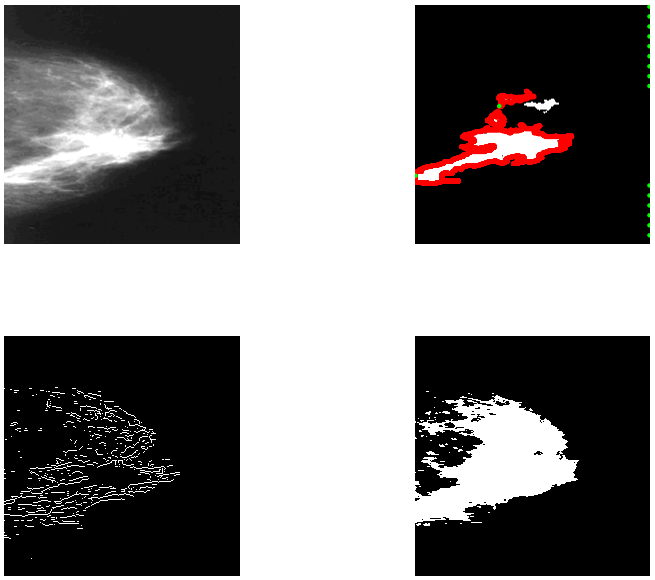

Fig. 4. Edge image segmentation: upper left: the original mammogram image, upper right: tracing the boundary of the ROIs, lower left: edge detection based Sobel operator, lower right: the binary segmented image

based on neural networks, then the files containing the extracted features are fed into the neural network which serves as classification method. For the purpose of classing, two images had been fed into the NN model, one class being classified as a malignant sample, and the other class as nonmalignant sample. On right positive, ANN is trained several times for the possible detection of B-cancer cells; after that, ANN can identify the specified ROIs if it belongs to class 1 or 0. The architecture of the ANN model commonly includes type of network, number of input and output neurons, transfer function, number of hidden layers as well as number of hidden neurons. Generally, the input and output neurons are problem specific. In this sudy, ANN is utilized in three layer feed-forward network with back-propagation algorithm. Therefore, the architecture of the network used in this study was as follows:

1. Number of hidden layers: 1 with 10 neurons in the hidden layer.
2. Number of inputs to the input layer: '$n$' (representing the number of texture features selected) depended on the model used.
3. Number of output layers: 2 (= number of classes)
4. The parameters of the network were as follows:

   - network: Feed forward back propagation
   - performance function: Mean Square Error (MSE)

- number of iterations: 1 000
- foal error = $10e{-}5$.

The detection method based ANN exhibits a high accuracy of correct categorization. All the trained and tested mammogram images were correctly classified using ANN for possible B-cancer detection. The network was trained 1 000 times based on the extracted features of the segmented objects using the feed-forward back-propagation network to minimize the MSE function. The improved results are attained by classifying the disease in the context of its features. Examples of class 1 containing suspicious regions classified by NNs as B-cancer are shown in Figures 5 and 6.

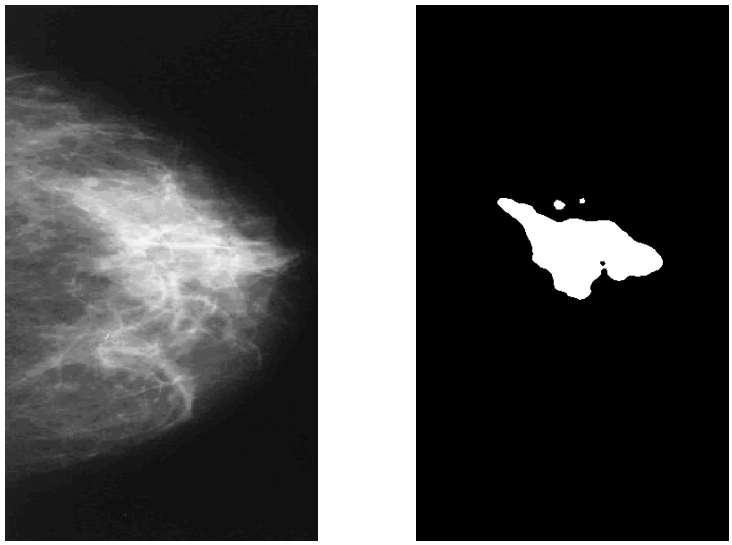

Fig. 5. Training case: A mammogram image containing B-cancer classified by ANN

Eventually, it can be seen from Figures 5 and 6 that ANN has the ability to recognize the mammogram images if the breast is infected with B-cancer or not. The convergence rate of classifying the B-cancer based ANN is recognized using the performance function. The MSE value of 0.0001 was used to measure the prediction accuracy; hence, the convergence rate achieved a high accuracy with 15 iterations. The convergence curve of the detection process is shown in Figure 7.

Figure 7 shows that the MSE value is converged to an optimum value, therefore, ANN has the ability to recognize the B-cancer in the mammogram images. This rate of detection indicates that the proposed method may be used as the basis for an effective prompting tool to assist radiologists in diagnosing B-cancer. The experimental results based the convergence curve indicate that the concept of automatic

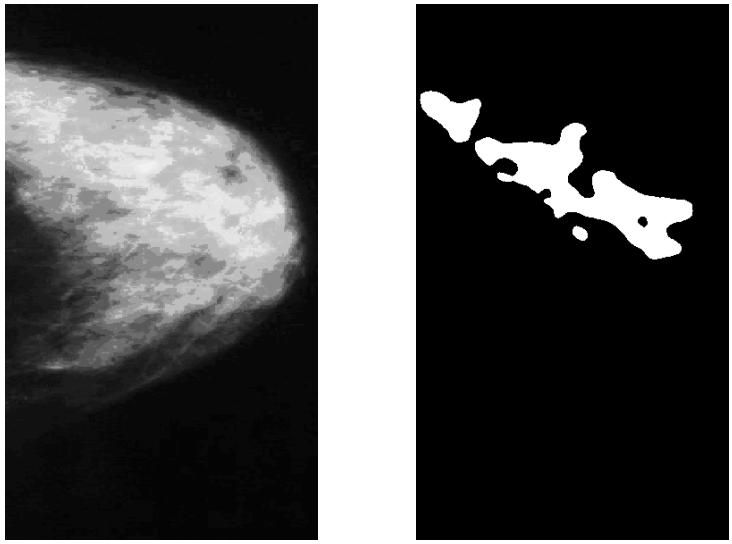

Fig. 6. Testing case: A mammogram image containing B-cancer classified by ANN

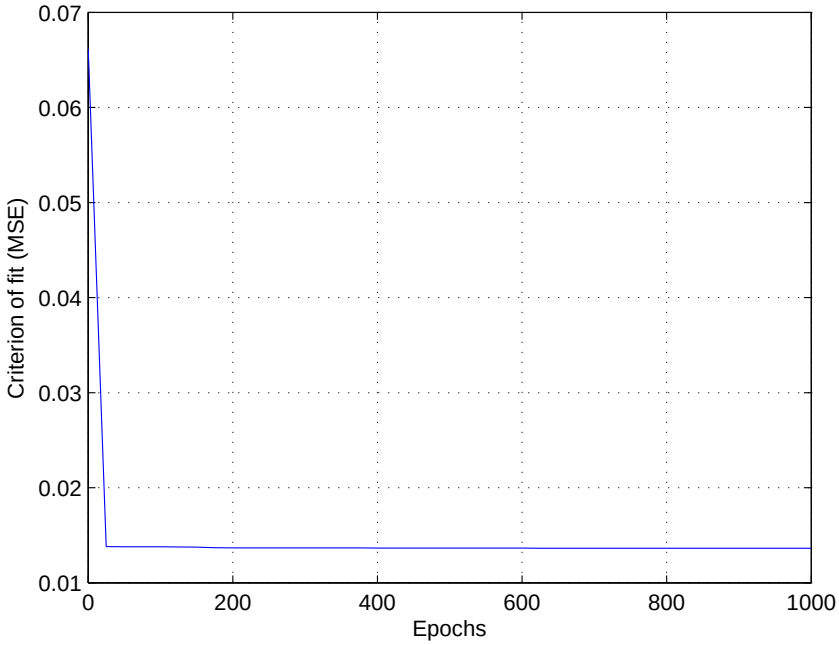


Fig. 7. The convergence curve of ANN

detection is very promising. That is, ANN is a powerful tool to model complex medical data which may capable of high classing accuracies.

## 8 EVALUATION CRITERIA

### 8.1 Recognition Rate

The classification accuracy as an evaluation metric is adequate provided that the class distribution of B-aner status is relatively balanced. The overall accuracy or reognition rate (Re_rate) is defined as the ratio between the total number of correctly classified instances and the test set size. Re_rate is defined as given in Equation (10).

$$Re\_rate = \frac{N_r}{N} \times 100\,\% \tag{10}$$

where $N_r$ is the number of correctly classified samples during the test run. $N$ is the complete number of test samples. Our test was performed for a total of 17 test samples, 16 samples are malignant and 1 sample is nonmalignant, and the same test was also performed for the trained samples; the rate of recognition for training and testing samples is shown in Table 1.

| Images set | Cancer ($N_{r1}$) | Non-cancer ($N_{r2}$) | Tot_No ($N$) | Reognition rate |
|---|---|---|---|---|
| Training set | 16 | 1 | 17 | 100 % |
| Testing set | 16 | 1 | 17 | 100 % |

Table 1. Recognition rate based on classification results

In Table 1 the reognition rate was computed using Equation (10), where $N_r = N_{r1} + N_{r2}$. It can be observed from Table 1 that the recognition rate is 100 %, hence, all training and testing mammogram images were classified correctly as class 1 or class 0. In the long run, the experimental results show the fact that this new approach can automatically identify the cancer cells. Thus, the screening experiments have shown that the concept of automatic detection was very convincing.

### 8.2 Sensitivity Versus Specificity

The importance of B-cancer segmentation and detection should not be depreciated. Increasing survival rates have been attributed to healthcare awareness campaigns and developing medical technologies. Regular examinations are keys to early B-cancer detection; however, as with any examination, there exists the possibility of misdiagnosis due to error. There are two classes of scientific error; systematic error and statistical error. A systematic error is the difference between what has been estimated of something and its actual state. For instance, human error is an example of systematic error which can be a factor in misdiagnosis. Also, a statistical error is

the difference between what has been computed of something and its actual state. Though, statistical errors are due to fluctuations in the measurement apparatus that are not predictable. Statistical errors are subdivided into two types of errors: Type 1 errors and Type 2 errors. A type I error is also known as a False Positive (FP). A FP occurs when a hypothesis states something is true when it is actually not true. For example, a patient receives a false alarm and is told she has cancer when she in fact does not. Type II errors are also referred to as False Negative (FN). A FN occurs when the status of something is reported as false when it is actually true. An example of this is when a malignant tumor is diagnosed as benign and the patient is told that she is clear of cancer when she actually is not. The following four relative metrics are calculated when comparing the classifier output of the constructed systems that were determined by biopsy:

**False Positive:** tumors which were not marked as tumor, but were classified as tumor.

**False Negative:** tumors which were marked as tumor, but which were not classified as tumor.

**True Positive (TP):** tumors which were not marked as malign, and that were also classified as malign.

**True Negative (TN):** tumors which were not marked as malign, and that were also not classified as malign.

The performance of such a diagnostic system is best described in term of its sensitivity and specificity quantifying their performance related to false positive and false negative instances. Sensitivity (SE) is the ratio of tumors which were marked and classified as tumor to all marked tumors. Thus, sensitivity is expressed as the ratio of number of true positives to the sum of true positives and false negatives; sensitivity is defined as given in Equation (11).

$$SE = \frac{TP}{TP + FN} \tag{11}$$

Specificity (SP) is the ratio of tumors which were not marked and also not classified as tumor to all unmarked tumors. Specificity is thus defined as the ratio between the number of true negative predictions to the sum of true negatives and false positives as given in Equation (12).

$$SP = \frac{TN}{TN + FP} \tag{12}$$

In reality, no inspection method can evaluate with perfect accuracy. Typically an instrument is considered to have an acceptable performance if its specificity and sensitivity are above 0.90. Table 2 displays a confusion matrix, which illustrates the relationship between the four possible values of an evaluation.

The results of a test performed with perfect sensitivity and specificity will all be either TP or TN and never FP or FN. We wrote a small program to evaluate our

| | Tumor marked malign | Tumor marked as **not** malign |
|---|---|---|
| Tumor classified as malign | True positives | False positives |
| Tumor classified as **not** malign | False negatives | True negatives |

Table 2. The confusion matrix. A "perfect" system detects with 100 % sensitivity and 100 % specificity.

metric criteria on 8 images; then, we tested our identified results with the actual results provided by radiologists who have high experience in diagnosing B-cancer. The data and the actual results are borrowed from Medical Hussein City with high efforts. Table 3 shows the confusion matrix for evaluation margin status in breast specimens, with sensitivity, specificity, Negative Predictive Value (NPV), and Positive Predictive Value (PPV).

| | | Status | | |
|---|---|---|---|---|
| | | Negative | Positive | |
| Histopathology | Negative | 15 | 1 | SP: 94 % |
| Margin status | Positive | 2 | 14 | SE: 88 % |
| | | NPV: 88 % | PPV: 93 % | |

Table 3. Confusion matrix for margin analysis on clinical specimens

The computed sensitivity (true positive (TP) detection rate) was 90.1 % with an average low false positive (FP) detection of 0.71 MCCs/image for the enhanced images using a modified k-fold validation error estimation technique. The corresponding computed sensitivity for the raw images was reduced to 81.4 % and with 0.59 FP's MCCs/image.

## 9 CONCLUSION AND FUTURE WORKS

In this paper an image-processing-based approach is proposed for breast cancer (B-cancer) segmentation and detection from digital mammogram images. The approach has been carried out with the assistance of morphological operations and Artificial Neural Networks (ANNs). This type of approach will help the diagnostics of B-cancer as a useful view. The proposed technique is composed of four main steps; in the first two steps the images at hand are enhanced and segmented using contrast stretching and a morphological approach, in the third step some texture features were computed, and finally, the extracted features were passed through a pre-trained neural network. Our experimental results indicate that the proposed approach can significantly support accurate and automatic detection of B-cancer disease. Based on our experiments, the developed neural network classifier that is based on statistical texture classification performs well and can successfully detect and classify malignant breasts with high precision. In future works we plan to use

support vector machine technique into the classification process for possible hiking up the efficiency and accuracy of the implemented tool.

## REFERENCES


[1] Peto, R.—Boreham, J.—Clarke, M.—Davies, C.—Beral, V.: UK and USA Breast Cancer Deaths Down 25 % in Year 2000 at Ages 20–69 Years. International Journal of Innovative Computing and Applications, Vol. 355, 2000, No. 2.

[2] Ranadhir, G.—Ghosh, M.—Yearwood, J.: A Modular Framework for Multi Category Feature Selection in Digital Mammography. Proceedings of the 12th European Symposium On Artificial Neural Networks – ESANN 2004, Brugges (Belgium) 2004, pp. 175–180.

[3] Elmore, J. G.—Wells, C. K.—Lee, C. H.—Howard, D. H.—Feinstein, D. R.: Variability in Radiologists Interpretations of Mammograms. The New England Journal of Medicine, Vol. 331, 1994, No. 22, p. 1493–1499.

[4] Thurfjell, E. L.—Lernevall, L. A.—Taube, A. A.: Benefit of Independent Double Reading in a Population-Based Mammography Screening Program. Radiology, Vol. 191, 1994, pp. 241–244.

[5] Murugavalli, S.—Rajamani, V.: A High Speed Parallel Fuzzy C-Mean Algorithm for Brain Tumor Segmentation. ICGST International Journal on Bioinformatics and Medical Engineering (BIME), Vol. 6, 2006, pp. 29–34.

[6] Alata, M.—Molhim, M.—Ramini, A.: Optimizing of Fuzzy C-Means Clustering Algorithm Using GA. Proceedings of World Academy of Science, Engineering, and Technology, Vol. 29, 2008, pp. 224–229.

[7] Vyborny, G. M.: Computer Vision and Artificial Intelligence in Mammography. Am. J. Roentgenol., Vol. 162, 1994, pp. 699–708.

[8] Kegelmeyer, W. J.: Computer Detection of Stellate Lesions in Mammograms. Proceedings of the SPIE Conference on Biomedical Image Processing and Three-Dimensional Microscopy, February 10–13 1992, San Jose, California, pp. 446–454.

[9] Gonzalez, R.—Woods, R.: Digital Image Processing. Prentice Hall, 3rd Edition 2008.

[10] Gulsrud, T. O.: Texture Analysis of Digital Mammograms. Ph. D. thesis, 2000.

[11] Pallikarakis, N.—Badea, C.: Medical Image Processing. Biomedical Technology Group, Dept. of Medical Physics. Current URL is `http://bme.med.upatras.gr/improc/enhancementpointprocessing`. See Topic: Contrast Stretching, Last accessed on May 23, 2008, 2008.

[12] Braik, M.—Sheta, A.—AYESH, A.: Particle Swarm Optimisation Enhancement Approach for Improving Image Quality. Int. J. Innovative Computing and Applications, Vol. 1, 2007, No. 2, pp. 138–145.

[13] Weyn, B.—van De Wouwer, G.—Daele, A.—Scheunders, P.—van Dyck, D.—van Marck, E.—Jacob, W.: Automated Breast Tumor Diagnosis and Grading Based on Wavelet Chromatin Texture Description. Cytometry, Vol. 33, 1998, No. 1, pp. 32–40.

[14] Doyle, S.—Hwang, M.—Shah, K.—Madabhushi, A.—Feldman, M.—Tomaszeweski, J.: Automated Grading of Prostate Cancer Using Architectural and Textural Image Features. ISBI 2007, pp. 1284–1287.

[15] Gonzalez, R.—Woods, R.: Digital Image Processing. Prentice Hall, second edition, 2002.

[16] Haralick, R. M.—Shanmugam, K.—Dinstein, I.: Textural Features for Image Classification. IEEE Transactions on Systems, Man and Cybernetics, Vol. 3, 1973, pp. 610–621,

[17] Pydipati, R.: Evaluation of Classifiers for Automatic Disease Detection in Citrus Leaves Using Machine Vision. Master's thesis, University of Florida 2004.

[18] Zhou, Z.—Jiang, Y.—Yang, Y. B.—Chen, S. F.: Lung Cancer Cell Identification Based on Artificial Neural Network Ensembles. Artificial Intelligence in Medicine, Vol. 24, 2002, No. 1, pp. 25–36.

[19] Scarpa, G.—Haindl, M.: Unsupervised Texture Segmentation by Spectral-Spatial-Independent Clustering. Proceedings of the 18th International Conference on Pattern Recognition 2006, pp. 151–154.

**Heba H. Al-Hiary** completed her B. Sc. degree in computer engineering in Computer Engineering Department, Faculty of Engineering Technology, Al-Balqa' Applied University in 2001. Later she received her M. Sc. degree in computer science at Information Technology Department, Al-Balqa' Applied University, Jordan in 2005. She has been working in a variety of image processing topics related to image restoration, segmentation, and compression. Her current work focuses on image processing, multimedia, data security and AI. She has published 7 papers (journals and conferences). Currently, she is working with the Computer Information System Department, Faculty of Information Technology, Al-Balqa Applied University, Al-Salt, Jordan.

**Basim Alhadidi** is presently an Associate Professor at Information Technology Department, Al-Balqa' Applied University, Jordan. He received his Ph. D. in Engineering Science in 2000 (computers, systems and networks). He received his M. Sc. in Engineering Science in 1996 (computer and intellectual systems and networks). He published numerous research papers in many topics such as computer networks, image processing, and artificial neural networks. He is a reviewer for several journals. He was appointed in many conferences as keynote speaker, reviewer, track chair and track co-chair.

**Malik Braik** received his B. Sc. degree in electrical engineering from Faculty of Engineering, Jordan University of Science and Technology, Jordan in 2000. Five years later, he received his M. Sc. in computer science at Department of Information Technology, Al-Balqa Applied University, Jordan. His research interests include evolutionary computation, image processing and computer vision. Currently he is a Ph. D. student at Electronic, Electrical and Computer Engineering (EECE) Department, University of Birmingham, U.K.